DOCUMENT RESUME

ED 108 510 FL 006 981

AUTHOR Keller, Howard H.
TITLE Polysemy and Homonymy: An Investigation of Word Forms and Concept Representation.
PUB DATE [74]
NOTE 43p.

EDRS PRICE MF-$0.76 HC-$1.95 PLUS POSTAGE
DESCRIPTORS *Computational Linguistics; Computer Programs; Information Processing; Language Instruction; Mathematical Linguistics; Second Language Learning; *Semantics; *Vocabulary; *Word Frequency; *Word Lists
IDENTIFIERS Homonyms; *Polysemes

ABSTRACT

Language teaching requires textbook material that contains the most frequent concepts of a language. The computer brings its tremendous information processing ability to the task of establishing word frequency rankings, but the computer is limited to counting word-forms and not semantic concepts. The most recent word frequency dictionaries, in fact, exclude parsing and lemmatization from their data base (Kucera and Francis, 1967: John B. Carroll, 1971). This paper describes the problems involved in adjusting a list of the 7,000 most frequent English words (word-forms) for polysemantic variants (e.g., cardinal "bird" vs. "church dignitary") and for homonyms (e.g., pawn "chess piece" vs. "pledge for a loan"). Polysemy and homonymy present a significant problem in that one word-form often expresses two or more differing concepts. The converse of this problem is synonymy--two or more word-forms expressing one concept (e.g., "freedom", "liberty"). The resolution of the difference between word-form and concept representation is important for accurate computerized frequency rankings and for concept inclusion in various "thousand" frequency groups. These problems will also be studied in connection with the establishment of a universal concept list for student review of foreign language vocabulary. (Author)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED108510

Polysemy and Homonymy: An Investigation of Word Forms and Concept Representation

Howard H. Keller[1]
Department of Foreign Languages
Murray State University
Murray, Kentucky 42071

FL006981

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Howard H. Keller
Murray State University
Murray, Kentucky 42071

The Problem

For decades the computer has been an invaluable aid in language research and in the preparation of language teaching materials. The complexity of grammatical rules and the large amounts of vocabulary have required automatic data processing routines for efficient handling. Vocabulary analysis is one major area of language pedagogy that is ideally suited for computer processing, and yet little work of any great range or significance has been done in this specialty.

The vocabulary system of a language poses a unique problem to the student of a foreign language. In order to communicate in a language a student must master several systems. He must learn some 30 to 45 sounds and their combinations, phonology, some 50 to 100 grammatical rules (and their ever-present exceptions!), morphology and syntax, and a set of vocabulary words that represent all areas of daily life which can be expected to occur in normal reading or conversation. The number and complexity of grammatical and phonological rules is significant, but this number is small in comparison with the size of a vocabulary that is required for ease in communication in that language.

Studies have indicated that an ability to recognize 7,000 words is sufficient to cover all areas of daily use, and an ability to use 3,000 words actively is a workable minimum for a person's expressive needs.[2] A statistical study of the 5,000,000 word corpus published in The American Heritage Word Frequency Book (AHWFB) shows that the most frequent 7,000 words will occur five times or more in every average group of 1,000,000 running words.[3]

Since a student must devote two or three years of study to mastering

a language, even this large number of vocabulary items is not a problem in itself. There is a twofold problem in the efficiency of the process, however. 1. The student encounters each of the 7,000 words in an unordered, random manner, and 2. He has no overview of how much he has learned and how much remains to be learned.

Except for series where related concepts are learned together (days of the week, months of the year, numbers, etc.), the student encounters each new word in a language text in a haphazard order. Even first year texts which claim to present the 1,000 or 2,000 most frequent words of a language differ widely in the actual lists of vocabulary words that they present to students.[4] A reading passage on HOUSE, for example, may have several words on FOOD, BODY, and HEALTH. A passage on MUSIC will probably not be limited to that topic, but could also include words from the categories of MIND, FEELING, and COMMUNICATION. It is a rare text that will then give a review list of vocabulary words arranged according to topics.

It is obvious that it is much easier to learn a word, e.g. Ger. Bucht 'bay' in a list of words that have a common topic, Ger. See, Meer, Fluss, Bucht, Hafen, Küste, Strand, rather than in a semantically unordered list, Ger. Buch, Buche, Büchse, Bucht, Buckel. In actual use of the language the student will see a word in a meaningful context; for this reason he should also have the benefit of a topical arrangement in his review. It is vastly more efficient and instructive to run one's eyes down a column of words that share a common semantic classification. In a list of unordered foreign words the learner is constantly shifting mental 'gears' as each new word calls to mind the vivid image of a new object: 'book, beech tree, tin can, bay, hump' (see the German example above).

This idea is stated in greater detail in the Wortfeld (word field) theory of Jost Trier and Leo Weisgerber. Their premise is that word contents or meanings of a form are rarely comprehended in isolation but rather are influenced and even determined by other words, and that one word evokes a picture of semantically related words ('field neighbors') in the consciousness of the speaker or hearer.[5] For example, in the series BIRD there is a lexical continuum of content, and a person who reads down the listing 'dove, pigeon, crow, raven, owl' would have a mental picture of one characteristic of each particular bird together with the general concept 'bird'. Each new word, 'parrot, ostrich, peacock, swan, stork', brings not only the characteristic of that bird with it but also a large element of the entire field as well. A rapid review of words from unrelated word fields in flash-card sequence would prove cumbersome due to the continuum of word field associations that accompany each word.

A student who ultimately hopes to recognize 7,000 words must have a view of his progress and an overview of the entire system. Many students who have diligently mastered 1,000 words at a beginning level feel that they already know a great proportion of the language, and so they are surprised when they continue to encounter commonly used words in lesson after lesson. A student becomes confused when he sees no recognizable end to the learning process, and when he has no way of really knowing which words he has already learned. For these reasons many students abandon language studies after one or two years with the idea that it is impossible to learn a language in a reasonable amount of time. Still more students never attempt a language because of a mistaken impression

of the awesome number of words that must be memorized as part of an alien 'code'.

The Topical Vocabulary Checklist

Depressed language enrollments dictate that a solution to this problem must be found. The key to providing order and system in vocabulary instruction is the division of 7,000 vocabulary items into manageable and workable categories and subcategories.

I propose a published checklist of all words which a student is likely to encounter in his language studies. This list will be divided into topical categories, and nouns will fill out and define the principal word groups. Verbs, adjectives, and adverbs will be listed under the appropriate noun categories. Each word entry will also carry a number from an authoritative word frequency dictionary indicating the frequency of that word. For most efficient use the list might be printed in several versions: a complete list of 7,000 words and a beginning and intermediate student's version of 2,000 and 4,000 words each. This topical checklist could be divided into 46 categories of differing lengths (see Table 1), and the list will be published in English. The checklist will serve as a type of preprinted notebook with a place for the student to write each new word as he encounters it either in his language classes or in his independent reading. There will be an alphabetical index to permit rapid word location or speedy transfer of large numbers of words from dictionaries. Since the list is in English it can be used uniformly for all European languages, and there will be blank spaces at the end of each topic for words not covered by the lists.

The best source of words for this checklist is Helen S. Eaton, An English-French-German-Spanish Word Frequency Dictionary.[6] This work is a composite of frequency studies in four languages, and provides fairly complete coverage of all topical specialties. Although it has the fault of age (which will be discussed later), it is a unique work and is widely available in paperback. It is a dictionary of meaningful words, and not a listing and counting of logical 'forms'. Its great advantage over word lists is that every word is parsed (annotated for part of speech), and lemmatized (separated for polysemy) by virtue of the translations of each word. The 6,500 words in the main part of the study can be divided into approximately 3,500 nouns, 1,500 verbs, 1,300 adjectives, and 200 adverbs. Nouns are the easiest to classify because they denote either concrete objects or easily comprehensible abstract concepts. Verbs, adjectives, and adverbs are listed with the appropriate noun category.

Creating a Concept List from Dictionary Word Listings

The computer is the ideal device for handling and classifying the 6,500 words in Eaton and ordering them into a complete topical vocabulary checklist. I would like to describe some of the problems in setting up this list, and I will outline the role of the computer in handling these problems.

The difficulties concerned with imposing workable categories on all useful words of daily life center around the general areas of word classification, word ordering, and word location. The computer was also helpful with the additional tasks of writing different versions of the list, indexing all formats of the list, and writing cross-references for the words or word

forms that appear more than once. Frequency notation for each word is also a problem, and the computer also permits simultaneous use of several frequency annotations obtained from very different sources.

The main considerations and procedures are the following:

1. Parsing, or the listing of each part of speech separately.

2. Establishing the topics and subtopics necessary to accomodate a corpus of 7,000 words.

3. Establishing a logical order of words within each topic and subtopic.

4. Assigning each word to an appropriate topic.

5. Polysemy: dividing one word with multiple meanings into several concepts, each with one specific meaning.

6. Synonymy: placing two different word forms of similar meaning under one concept listing.

7. Establishing concept listings with English words that will still permit the listing of foreign words with different definition extensions.

1. Parts of speech must indeed be listed separately. A one letter code can be used to mark words for their part of speech (N, V, A, D for noun, verb, adjective, and adverb), and the sort on these letters is always done first. It will be shown later during a consideration of computerized lists that a great many word forms in English can occur simultaneously as nouns, verbs, or adjectives.

2. A series of topics and subtopics must be established to provide a logical and useful division of reality. Philosophers and scientists have proposed and reworked divisions for reality long before the advent ofthe

categories of Aristotle, and this process of classification has continued to the present day. The three criteria of an acceptable vocabulary organization are that it be complete, that the groupings be plausible and logical, and that the 7,000 basic words of a language inform the various categories and subcategories in a balanced manner. A glance at Table 1 gives an indication of the reasonableness of its subdivisions, and a view of the finished product will determine if all 7,000 words have been assigned in even proportions.

3. Every word in the study must be part of a logical order under its appropriate topic. Many arbitrary decisions must be made in establishing this order, since there is obviously no self-evident way of classifying reality. It is wise to establish a limit to the number of words in a category (40 has proven to be a convenient number). Since subtopics will usually contain less than 40 words, it becomes easy for the user of the list to locate the word he is looking for once he has found the correct topic under which it is classified. Please see Table 2 for an illustration of a sample topic. In examining these lists the advantages of a topical listing over an alphabetical listing become immediately evident.

Each of the 7,000 words in the study has been keypunched on an 80 column data card. This is sufficient to allow a full statement of a 'concept' of several synonyms, the punching of several frequency annotations, and an eight place topic and order code. All cards were rearranged to establish category location and an internal logical order for each category. A code number was then given to each word so that a sort on this number would print the lists out in the desired logical order. A three digit number was sufficient

to prepare each category for internal ordering, since no category has more than 250 word-concepts. Since words are constantly being reshuffled and reassigned to get the most consistent arrangement, two additional digits were reserved for the code number to permit insertion of new words between any two consecutive words in the list without changing the code numbers of all remaining words. A decimal point was inserted as a boundary between the two digit category code (01 to 46) and the five digit subcategory (internal ordering) code.

Subcategory divisions and headings are indicated by a change in the first digit following the decimal point. Thus, 10.100 to 10.199 is assigned to TREE, 10.200 to 10.299 is for PLANT, and 10.300 to 10.399 is for FLOWER. The printing routine causes a line to be skipped each time the index number indicates a new subcategory. This ensures a convenient visual separation within each topic. The additional two digits do not print out unless a word has been inserted between two previously coded words, but these digits give the list the potential of great flexibility.

4. Many words have presented problems in assigning them to one unequivocally unambiguous category. A large number of words can indeed fit with equal justification under two or more categories. A glance through the following examples gives some idea of the problem: doctor: PROFESSION (12) or HOSPITAL (22); chicken, turkey: ANIMAL (06), BIRD (07), or FOOD (17); lamb: ANIMAL (06) or FOOD (17); invention, discovery: TECHNIQUE (13) or SCIENCE (27); theory: MIND-INTELLECT (24) or SCIENCE (27). Abstract nouns such as 'operation, building (act of), dissolution, motion, return, fall (tumble), hold, burning, fact, manner, process, means, advantage,

change, connection' were particularly hard to place in one exclusive category.

For this reason the computer renders a great service in permitting unlimited word shufflings until the best internal consistency has been reached. This process still leaves arbitrary decisions, and so a computer-generated index is of great value in locating words that do not have an obvious location. It is indeed tempting to list a word two or three times if it fits two or three categories, but this destroys the one-to-one correspondence between the topical vocabulary checklist and the source lists. It will be seen later that multiple listings of single word-forms will be required when one word-form expresses several meanings or concepts.

5. The phenomenon of polysemy is probably the greatest obstacle in what might otherwise be a mechanical process of taking a finite list of 7,000 words and simply rearranging them according to meaning similarities.

Polysemy is defined as the fact that a word-form may have more than one meaning or may designate more than one object or concept in the world outside language. An extensive example is furnished by the word-form 'head' which can designate objects or concepts as diverse as: head of a body; mind (a good head); a drug head (user); the obverse (heads) side of a coin; an individual within a group (count heads); an animal within a herd (head of cattle); a boss, leader, chief, or director (head) of a department; the pressure of liquid or vapor (head of steam); the foam on an effervescent liquid (head of beer); the tip (head) of an abscess, boil, or pimple; a turning point or crisis (come to a head); the head of a screw, of a bolt, of a pin; the head of a hammer; the recording head of a tape deck; a head of

cabbage, lettuce, or cauliflower; the head of a door or window; the head of a bay; the head of a bed, table, drum, ship (front part), of a cape or cliff; of a book or page; a topic or category; and finally, a ship's head (bathroom).

The example is perhaps an extreme one, but it illustrates how many different figurative concepts can be named from one literal meaning of "head: topmost part or most important part of a larger body." These different meanings do indeed merit individual consideration because they are not predictable from the source concept 'top of a body.' We see proof of this in the fact that various foreign languages translate most of the figurative extensions of 'head' with a variety of word forms that come from different figurative analogies and are formed from different roots.[7]

It also becomes apparent from reading through a full list of definitions for 'head' that it is impossible to establish a precise number of meanings or draw well defined boundary lines between the various specialized concepts. Many technical meanings occur in very limited environments, and once the connection with the concept 'head' is established, the analogy becomes evident. Is the head of a bed qualitatively different from the head of a door or window? A hammer and a pin are very different tools, and yet the concept 'head' seems appropriate for the elongated projections of both these objects.

Polysemy is a feature of a great number of words. A large percentage of the basic words in a language have dictionary entries with several numbers to indicate the multiple concepts or specialized definitions. The range of these analogies may be either narrow, where the figurative link is obvious ('head' of bed, window, door, or table), or wide, where rather diverse

objects are gathered under one word-form ('solution': solving a problem vs. dissolving an element in water; 'staff': rod or stick vs. group of executive assistants.)

The difference between a homonym (a member of a set of two or more words that have the same sound and often the same spelling but differ in meaning) and a word with multiple meanings is that the semantic link is always present in polysemy but not in homonymy. 'Pawn' in chess and 'pawn' as a pledge for a loan are a homonym pair since they do not share a common etymology. It is especially true in examples of wide range polysemy that the semantic link may not be evident at first ('cardinal': bird vs. church dignitary), but the "bridge" either becomes apparent after some examination or it can be established in a good reference dictionary.

The computer offers useful assistance at this point, because a topical vocabulary checklist must be a compilation of concepts, not of word-forms. The fact that 'cardinal' is listed under RELIGION in the order 'clergy, pastor, priest, bishop, archbishop, cardinal, pope' must not exclude a listing under BIRD in the order 'oriole, blackbird, cardinal, finch, sparrow.' The computer enables the compiler to keep track of the source words with multiple meanings after these word-forms are listed in separate categories, and the computer also permits the compilation of a separate list of polysematic word-forms in the appendix to the checklist.

This list is of interest to language students because it shows how a given language makes ample use of figure and metaphor to supply linguistic symbols or "code forms" for the thousands of objects and concepts that

exist outside that language. If the word equivalents in a target foreign language are also keypunched and are matched with their equivalents in the computerized English corpus, and if word-forms in the foreign language are also coded for multiple meanings, the computer can repeat the process of compiling an appendix of polysemantic words in the target language. Language students of all levels can then gain an insight into the process of concept-representation and concept-adaptation in the target language, and the students will see how a speech community exploits its native word stock to cope with the communications needs of a changing culture.

The compiler of a concept list must use judgement in selecting the most common denotations of a word--those that are relevant to daily use-- and he must bypass highly specialized or technical denotations. In most cases a foreign language will translate the various concepts of a specific word-form with different words, and so one can examine the frequencies of these words in the foreign language to determine the value of the extended or technical meanings.

Table 3 gives a small indication of the forms that required multiple listings because of polysematic meaning diversity. The most general trend concerns the grouping of one concrete image and several additional abstract or intangible denotations: e.g. 'way': road and method; 'summit, peak': mountain and any high point or climax; 'genius': the person and the quality of mind. Still other groupings concern a person or thing and a cognate activity:'gossip' groundless rumor and a person who gossips; 'residence': a dwelling and the act or a period of living somewhere; 'building': a structure for shelter and the act of

constructing; 'painting': a portrait or painted picture and the act of painting anything;'entrance': the door itself and the act of going into something. A limited number of polysematic pairs also arise from the need for technical terms that are restricted in use in a particular specialty. Chess and card-playing terminology provide a good example in the series of meanings of the word-forms: 'king, queen, knight, bishop' or 'king, queen, jack, spade, heart, diamond, club.' These two specialties have in turn added some of their own technical terms to the general language use: 'pawn': a person used as an object by another person; 'ace': a person who is an expert in his field.

Any attempt to write down one language's categorization of reality must take into account the fact that each language is in a constant state of change. Every language is continually adding new words to its vocabulary inventory, and is continually changing the meaning of older words. It is a delight to see this process at work,because languages prefer to give a form to a new object or concept from an existing root by using analogy, poetic description, or some other figurative device. It is unusual for a language to build new word-forms from previously unused sound combinations in the manner of jabberwocky.[8] After the language user becomes aware of this process, he can note how a language not only supplies sounds and sound combinations for extralinguistic objects, but also makes its own descriptive statements about these objects and their place in the world. We are so accustomed to speaking our native language that we no longer notice the underlying metaphor in thousands of words. One of the rich experiences in learning a foreign language is the gaining of a fresh view of this

vigorous process by using a set of entirely different meaningful word-forms and word roots.

To summarize the problems of word listing, topic location, and polysemy: where a word-concept can occur under two or more categories, a decision must be made for one category, since only one concept is involved and only one listing is possible. Where one word-form denotes two or more basic concepts, one must place the several concepts under their appropriate topics, and not restrict the listing to the one word-form.

6. The converse to the problem of polysemy (one form and several concepts) is the problem of synonymy (several forms that express only one concept). John Lyons and many other authors point out that no two synonyms have exactly identical (coterminous) meaning extensions nor are mutually substitutable in every context or meaning environment.[8a] Yet English is a language that is particularly rich in synonyms, and European languages often have only one word to translate many close synonym pairs in English. In order to establish an economical concept list that will not involve repetitious listing of foreign words to accomodate series of English synonyms, many decisions must be made in word formats. For example, the synonym pair 'freedom, liberty' could be listed as one entry separated by a comma, or the two words could be listed as two discrete entries 'freedom' and then 'liberty.' Automatic data processing procedures are useful in resolving this problem, since they permit the indexing of all words that appear not as a main entry but as a synonym after a comma. A special code symbol is enough to mark

these words as secondary synonyms.

Table 4 gives a short indication of some of the more common synonym pairs. A glance at this list will indicate that the degree of difference in some pairs is greater than in other pairs. In many pairs the difference is a technical one, and in loose speech one member of the pair might be substituted for the other member, even though there is indeed a difference: e.g. tortoise and turtle, hare and rabbit, alligator and crocodile, etc. Since the goal of the vocabulary checklist is to classify the more basic concepts of a foreign language that a student might encounter, even a loose grouping of synonyms (where each member of the pair has an additional meaning of its own) is permissable.

The only complication created by synonym pairs is that it is difficult to transpose a frequency count of an individual form to that of an individual meaning-concept. If we wish to establish the frequency rank of the concept 'pigeon/dove' where the form 'pigeon' has a frequency-per-million (FPM) count of 7.1511 and the form 'dove' an FPM of 4.1270,[9] we must decide whether to use the higher individual FPM (of 'pigeon') giving us a ranking in the 6,300 frequency group, or to add the values of FPM for the two forms 'pigeon' and 'dove' giving us a total FPM of 11.2781 for the concept 'pigeon/dove' and a consequent ranking in the 4,800 frequency group. For many words or concepts the <u>total</u> FPM becomes important in establishing criteria of inclusion in frequency rankings (thousand groups) and in lists defined by these frequency rankings. If our list is limited to the first 7,000 words of a language, neither the form 'poultry' nor the form 'fowl' would merit inclusion due to their low FPM and consequent low rankings

('poultry' 4.7946, 8,000 group, and 'fowl' 3.1764, 9,900 group). The single concept 'poultry/fowl' would merit inclusion, however, because the total FPM of the two word-forms, 7.9710, places it in a ranking with the 5,900 word group.

7. The last major problem in establishing a concept-oriented topical vocabulary checklist arises from the fact that each language has differing extensions in the denotations of its basic word-concepts. Some graphic examples of this are Rus. ruka for both 'hand' and 'arm'; Rus. noga for both 'foot' and 'leg'; Fr. doigt for both 'finger' and 'toe'. Very often the sum of a set of words from two different cultures will have the same total extension or denotation, but the subdivisions of the set will have different proportions in each language. For example, many European and American cultures have three mealtimes in a day, but the definition of the individual members of the set 'breakfast-lunch-dinner-supper' varies extensively in terms of size of the meal, preparation (hot or cold), and time of serving. All European cultures use a 24-hour day, and yet the times and lengths of the following subdivisions vary from language to language: morning, afternoon, evening, night. In fact, Russian does not have a commonly accepted form for 'afternoon', and while French and German do have a word for 'afternoon' (après-midi, Nachmittag), they have no equivalent for the greeting 'Good Afternoon'. In Spanish, one word-form is often used for both 'afternoon' and 'evening' (tarde), and the word-form noche can be used for both 'evening' and 'night'.

D. A. Wilkins in his interesting work, Linguistics in Language Teaching, exaggerates the problem when he says: "The physical world does

not consist in classes of things, nor are there universal concepts for each of which every language has its own sets of labels. Language learning, therefore, cannot be just a matter of learning to substitute a new set of labels for the familiar ones of the mother tongue. It is not difficult to find a word of equivalent meaning in a given linguistic and social context. It is most unlikely that the same word would prove equivalent in all contexts. Every language classifies physical reality in its own way." [10]

There is indeed a good deal of truth in this statement, but one must not interpret it to mean that it is impossible to translate language A into language B because every concept in A is not coextensive with every concept in B, and every word-form in A does not have an exact equivalent in B. Even though Wilkins's statement is basically true, we can still translate two different languages with a fair degree of approximation, and we can also set up concept surveys of those objects in extralinguistic reality which any two European languages might reasonably be expected to express.

If the English words selected to represent concept groups are kept general, and if the user of the list is prepared to accept blank spaces for some concepts (Rus. 'afternoon', Ger. 'efficient, frustrated') and the necessity to enter several foreign words for other single concepts ('cousin': Ger. der Vetter, die Cousine; 'return': Fr. revenir, retourner, rentrer, renvoyer; 'box (container)': Ger. die Schachtel, die Schatulle, der Kasten, die Kiste, das Kästchen, das Etui, das Futteral, der Karton, der Koffer, die Dose, die Büchse), the user will see that the English word-forms for these concepts are neither absolute nor universal.

All useful sets must appear in the topical list, but the component members of these sets should be stated in a general manner, since foreign languages rarely have equivalents for all members of a set of objects. An example of such a series or set is ROAD, where not every foreign culture can match each individual member of the set: 'path, alley, lane, way, street, road, avenue, boulevard, highway, expressway/freeway/Interstate'. Different cultures have differing subdivisions for COAT/JACKET as suggested by the English concept-divisions: 'overcoat, topcoat, frock coat, coat, three-quarter length coat (car coat), field jacket, lumber jacket, windbreaker/anorak, suit jacket, suit coat, C. P. O. shirt,' etc. This concept list is overly specific. A series list is more useful when it has more general subdivisions ('overcoat, coat, field jacket, suitcoat). It is also helpful if the user is made aware of the value of giving cultural correspondences or equivalents, rather than making exact translations.

The Advantages and Limitations of English Word-Forms for Universal Concepts

There is a great advantage in writing the concept list in one language only. The student receives active practice in using the list by writing down each new foreign word as he learns it. This is much more useful than the student's running his eyes down an already prepared list of equivalents in both his native and foreign languages. The concept of the topical vocabulary checklist as a prepr[illegible]ed notebook was mentioned earlier. An additional use of an "English only" list is for students who are learning more than one foreign language. If that student has two blank copies of the checklist, he has a very graphic device for measuring and matching his knowledge in

both languages, and he can transfer those insights gained from studying his first foreign language into his second foreign language. When these two foreign languages are closely related (e.g. French and Italian, Spanish and Portugese, German and Danish, Russian and Polish), the user of the lists can see the great similarity of vocabulary, and can build his knowledge of the second language on his knowledge of the first.

The only real disadvantage of an "English only" concept list is that it will not have a specific entry for the many culturally unique aspects (food, clothing, etc.) of a foreign culture. Here again, judicious amounts of blank space after each subheading will permit the student to catalog culturally unique words such as Sp. tamale, taco, enchillada, chili; Rus. vodka, borščč, ščči, kolbasy, zakuski, etc. No concept list or implied division of reality could ever pretend to be complete. Blank spaces give the checklist another dimension of flexibility, and the student is also free to "personalize" his list by expanding those areas where he has a specialized interest in technical words that do not appear in the first 7,000 word frequency range.

A good topical vocabulary checklist will stand or fail on the strength of its ability to classify, order, and list words. The main section of this article has shown the seven problem areas where ADP routines can help to achieve a balanced list. The last consideration of such a list are the frequency notations of the words themselves.

The Need for Frequency Notations

The list is based on the 7,000 most frequent words of four languages, but it is still important that each word have an individual frequency notation. This is a great help to the student in that it eases the arduous process of learning and reviewing large numbers of foreign vocabulary words. The shorter versions of the list mentioned earlier (2,000 words and 4,000 words) are a morale booster in that they establish intermediate stages of proficiency--subgoals that give the student a sense of accomplishment when he masters them. If a beginning language student sees the full array of 7,000 language concepts that remain to be learned, he could be easily overwhelmed by this vast mass of words, and could decide to discontinue further language study.

The computer permits rapid establishment of intermediate listings under any frequency criteria that a compiler or publisher might desire. The attendant task of creating the various indexes for each intermediate version is also made easy by computer routines.

One major reason for any topical listing is to reduce large categories to manageable levels. If each word has its own frequency"identity",the student can set intermediate goals on his own even if he is using the full 7,000 word version of the list. He can include or exclude subcategories according to his interest in them or his need for them, and he will no longer be overawed by vast quantities of words that he must learn. Even though the student's intuition or common sense will tell him that some words are more important than others; the student still needs frequency indications in the majority of topics to gain perspective concerning the importance of

large numbers of vocabulary words. It will be obvious to the student that the names of berries, grains, trees, and birds are not of high frequency, and that 'eyelid, eyebrow, temple, nostril, knuckle, knee, ankle' are definitely less important than 'head, eye, nose, finger, leg, foot.' It will be very difficult for the student, however, to determine which abstract nouns are more important than others, and which verbs and adjectives should be learned before others.

A vocabulary notation next to each word gives the student an additional dimension of choice in learning words. He can establish different limits in each topic according to which thousand group he wishes to include. He can learn all words in categories of special interest to him, and he can omit, for example, all words beyond the first or second thousand group in other categories. If he must stop his language studies at any time, he has an exact record of the categories and frequency limits he has studied, and he will not need to duplicate these words when he resumes his studies at a later date.

Source of Frequency Notations

As mentioned earlier, the use of Eaton's four-language frequency dictionary as a source list brings the advantage of giving the 6,500 most common words of these languages together with a frequency annotation for each word. The original date of the work is 1940, and it in turn is based on four monolingual frequency dictionaries that date from 1932, 1929, 1927, and Kaeding's monumental word count of 1898.[11] Thus, many words in Eaton are "dated", and there are many modern words which are not included.

Eaton's frequency dictionary has three valuable advantages that more than compensate for its age: 1. every word is marked for part of speech; 2. every word is implicitly lemmatized into its major (most frequent) concepts; and 3. every word is listed with an equivalent in three other languages.

The two most useful modern word studies are based on the Kucera and Francis count of the Brown University Corpus (BUC) [12] and John B. Carroll's count of the American Heritage Intermediate Corpus (AHIC).[13] The BUC is a computer analysis of 1,000,000 running words of all specialized topics (sciences, humanities, literature, etc.) and areas of difficulty, and was keypunched in 1963-64. The AHIC is a computer count of 5,000,000 running words of graded school reading material from grades 3 to 9. The AHIC was designed to produce a citation base for The American Heritage School Dictionary, and has also been used in the preparation of vocabulary questions on college entrance examinations.

It must be emphasized that these two studies are not word counts but rather counts of logical forms. The studies define a 'word' as any string of graphic characters preceeded and followed by a space. This logic, while necessary for computational analysis, renders the BUC and the AHIC of little use in setting up frequency numbers for meaningful concepts.

The lack of parsing (indication of part of speech) nullifies the value of both studies for word analysis. All noun plurals, noun possessives, 3rd sg. verb forms, and past tense or past participle verb forms are counted separately from their non-inflected noun and verb forms. An additional complication is that all nouns, verbsm and adjectives with the same

graphic form (e.g. 'blue, gray, counterfeit, double, upset') are listed as one form only, with no record of the part of speech used in the source texts. There are hundreds of English words in the first 7,000 frequency range that have an indentical form for either noun/verb (e.g. 'want, use, work, look, talk, help'), noun/adjective (e.g. 'drunk, cold, noble, concrete, secret, magic'), or verb/adjective (e.g. 'dull, blunt, smoothe, equal, blind, clean'). To cite the word 'word' as an example: the AHWFB lists seven graphic forms: 'word, Word, word's, worded, wording, words, Words'. In order to find the total frequency of the concept 'word', one must add all individual frequencies of the seven forms. Even here there is still no accurate indication of the relative values of 'word' NOUN and 'word' VERB.

The BUC and the AHIC can provide a useful service for the topical vocabulary checklist, however, by providing a "backup" set of frequencies for the words listed in Eaton. A check of the first 7,000 BUC and AHIC listings will at least indicate which "modern" words are missing in Eaton, and which words have gone out of frequent use since the compilation of the Eaton source frequency dictionaries. The BUC and AHIC frequency notations would not need to be published with the topical vocabulary checklist, since they would be of interest only to the language researcher. A language scholar would indeed find fascination in an examination of the shifting values of the most frequent 7,000 words of a language in a period of 30 years.

Conclusion

The hurdles posed by polysemy and synonymy are not insurmountable in the establishment of a concept-oriented topical vocabulary checklist. The computer provides a vital service in enabling the compilation of many preliminary versions. Automatic data processing permits countless word and category reshufflings until the word arrangement seems "just right", and until the list shows a high degree of internal consistency.

Even though it is impossible to classify and divide reality into a fixed list of universal categories and subcategories, the attempt to do so produces a list that has great merit as a teaching device. Even the discrepencies between the English list and the target language serve to give the student an insight into how both languages order and express reality.

As the student uses the list, he will gain respect for its completeness, and he will see that the vast complexity of reality is much more easily comprehended in a foreign language if the several thousand word forms are rearranged into workable categories. Even if the student encounters new foreign vocabulary items in an unordered or "flash card" sequence, he can write them down under ordered topics and he can compare them to the semantically related words in those topics that he already knows.

The checklist in either a complete or an intermediate version shows the student a visible end to the process of mastering a foreign language, and it permits the student to monitor his progress in attaining fluency in that language. At all times the student has a clear view of how much

he knows and how much remains to be learned. This perspective is essential, since otherwise the learning process would be an exercise in learning 7,000 unordered words.

The topical checklist is a useful device for the many necessary vocabulary reviews that are associated with language learning. As a student checks off words that he knows in a category, he receives encouragement to learn the remaining words, especially when the proportions indicate that he has already learned 70% or 80% of the words in a given topic.

If a topical vocabulary checklist is made a part of the curriculum of a language course, the student will receive an added awareness of his mastery of the language: he will know what he knows, and he will know what he doesn't know. A student with a feeling for his own vocabulary strengths and weaknesses will be much better able to participate in speaking and conversation exercises, since he will know those areas where he has a detailed vocabulary competency, and those areas where he must paraphrase his thoughts in order to be understood.

Awareness of his progress in language is a valuable encouragement to students of any level in any language course. Today's declining language enrollments place a great deal of value in encouraging students in the face of the complexity of a foreign language.

Notes

[1] This study has been made possible through a grant from the Murray State University Committee for Institutional Research. The grant will finance the computer work necessary to prepare a topical vocabulary checklist. I wish to thank the committee and its chairman, Dr. Kenneth E. Harrell, for the faith they have shown in this project.

[2] Robert Lado, Language Teaching: A Scientific Approach (New York: McGraw-Hill, 1964), p. 117, gives these figures as a result of examining various general service word lists.

[3] John B. Carroll et al., The American Heritage Word Frequency Book (New York: American Heritage Publishing Co., 1971).

[4] This is due to great differences in frequency lists which in turn are compiled with differing criteria from different types of source texts. A study of the five leading Russian textbooks and the three accepted word frequency lists showed a great variation of inclusion and exclusion (Nicholas P. Vakar, "Statistical Methods in the Analysis of Russian," Slavic and East European Journal, 11 (1967), 59-65.

[5] See the entry "Feld" in Werner Welte, Moderne Linguistik: Terminologie, Bibliographie. 2 Bde. (München: Max Hueber Verlag, 1974), for a concise treatment of the word field theory and for a useful list of descriptive literature.

[6] Helen S. Eaton, An English-French-German-Spanish Word Frequency Dictionary (New York: Dover Publications, 1961). This work is a reprint of an earlier edition entitled Semantic Frequency List for English, French, German, and Spanish (Chicago: University of Chicago Press, 1940). This dictionary establishes the frequency ranking of a word by "thousand groups." Eaton averages the frequency annotation for each word in each of four languages. Thus, the notation 1.0 designates a word that occurs in the first thousand group in all four languages. A notation of 2.6 indicates an average frequency ranking of 2,600 in all four source dictionaries; 3.9 indicates 3,900, and so on.

The actual number of word-concepts in Eaton is 6,473, but I use the number 7,000 in various sections of this article because it is a convenient round number and is approximated by additional proper nouns listed in the appendix of the Eaton dictionary.

[7] W. F. Mackey, Language Teaching Analysis (London: Longmans, 1965), p. 69, gives the same example in more detail:

head	of a person	tête
	of a bed	chevet
	of a coin	face
	of a cane	pomme
	of a match	bout
	of a table	haut bout
	of an organization	directeur
	on beer	mousse
	title on a page	rubrique

[8] The 5,000,000 word corpus of the AHWFB contains a number of nonsense words--sound combinations (or word-forms) that would be acceptable according to the canonical shape limitations of English, but that do not happen to have a concept-meaning assigned to them: billitch, binning, brendly,

brosket, clob, crobble, floffle, frobish, froon, frumious, frums, gloobed, gorbed, gribble, grop, grumple, motch, nugful, quinking.

[9] The AHWFB goves two listings for each word type in its 5,000,000 word corpus. The frequency (F) listing is the simple record of how many times that a specific word-form has occured in the corpus, from 'the' with an F of 373,123 to the 35,079 word-types which occured only one time (hapax legomenon). The absolute frequency notation (F) has been adjusted for even distribution throughout all 17 text specialties, however, by the calculation of a distribution value (D). The result is a second frequency indication, a theoretical but statistically plausible frequency-per-million value (FPM). Thus, the form 'but' has a fairly proportional distribution throughout the 17 text specialties, and has a FPM value of 73,122.8. The form 'cowherd' has an F of 5, but is limited to only one specialty and has a FPM value of 0.2284.

The less frequent FPM values are usually calculated to four places after the decimal point, and the FPM represents the statistical probability of encountering a specific word-form in a corpus of 1,000,000 words. The FPM value has an additional advantage of permitting comparison with the many word frequency studies which have been done on corpora of 1,000,000 running words.

[10] D. A. Wilkins, Linguistics in Language Teaching (Cambridge, Mass.: The MIT Press, 1972), p. 119.

[11] The source dictionaries are: Edward L. Thorndike, Teacher's Word Book of 20,000 Words (New York: Columbia University Teacher's College, 1932) which uses a 9,565,000 word corpus; George E. Vander Beke, French Word Book (New York: Macmillan, 1929), a listing of 6,067 word-types from 1,000,000 running words; Milton A. Buchanan, Graded Spanish Word Book (Toronto: Univ. of Toronto Press, 1927), a listing of 6,702 word-types from 1,200,000 running words; and Friedrich W. Kaeding, Häufigkeitswörterbuch der deutschen Sprache (Berlin: E. S. Mittler und Sohn, 1898), a listing of 79,716 word-types from 10,910,777 running words.

[12] The 2,000 500-word texts that make up the BUC are available on computer tape. The count exists in published form: Henry Kučera and W. Nelson Francis, Computational Analysis of Present-Day American English (Providence: Brown Univ. Press, 1967).

[13] The American Heritage Intermediate Corpus count has been published as The American Heritage Word Frequency Book (op. cit.).

[8a] John Lyons, Introduction to Theoretical Linguistics (Cambridge: Cambridge Univ. Press, 1968), p. 447-48. See also Lyons, Structural Semantics (Oxford: Blackwell, 1963); Stephen Ullman, The Principles of Semantics (Oxford: Blackwell, 1957); p. 108-09, and Ullman, Semantics: An Introduction to the Science of Meaning (Oxford: Blackwell, 1962), for a fuller treatment of this idea.

Table 1

The headword for each category is given in one or two main words. In this overview the headword does not give a complete or detailed idea of every word in that particular category. The number of nouns in each category is indicated in order to give an approximate idea of the relative size of each category. It must also be remembered that these numbers are subject to change, since words are still being reshuffled among categories.

Category 1, 139 nouns.
world
the country
mountain
water

Category 2, 35 nouns.
agriculture
crops

Category 3, 63 nouns.
city
store, shop
castle, palace

Category 4, 29 nouns.
weather

Category 5, 70 nouns.
element, mineral
metal
wood, lumber
element
color

Category 6, 65 nouns.
animal (domestic)
animal (wild)

Category 7, 38 nouns.
bird

Category 8, 10 nouns.
fish

Category 9, 18 nouns.
insect

Category 10, 79 words.
tree
plant
flower

Category 11, 49 words.
time
day
calendar

Category 12, 69 words.
profession,
occupation

Category 13, 71 words.
administration
management
technique

Category 14, 107 words.
person
friend
fool, idiot

Category 15, 45 words.
family

Category 16, 33 words.
life

Category 17, 138 words.
food, meal
bread
meat
vegetable
fruit
dessert
drink

Category 18, 138 words.
dwelling,
residence
house
living room
dining room
kitchen
bedroom
bathroom

Category 19, 82 words
machine
tool, implement
box, container

Category 20, 109 words.
clothes
suit
dress
cloth
sewing
jewelry
costume

Category 21, 109 words.
body
head
trunk
limbs
organ

Category 22, 51 words.
sickness
hospital

Table 1 (contd.)

Category 23, 30 words.
luck, chance
success
failure
misfortune

Category 24, 40 words.
mind, intellect

Category 25, 244 words.
feeling, sense
mood (happiness, anger)
love
care, worry
sorrow
honor, respect
hate, hatred

Category 26, 37 words.
education
school

Category 27, 24 words.
science

Category 28, 104 words.
culture
painting
sculpture
theater
music
musical instrument

Category 29, 25 words.
literature

Category 30, 45 words.
language

Category 31, 122 words.
communication
conversation
truth, lie
answer
discussion
sign, symbol

Category 32, 83 words.
writing
book
newspaper
letter

Category 33, 142 words.
religion
church
worship, rel. service
clergy
marriage
funeral
superstition

Category 34, 33 words.
company, society
way of life, habit, custom

Category 35, 53 words.
club, association
hospitality
smile, laugh
gift, present

Category 36, 61 words.
hobby
toy
sport
riding
hunting
cards
chess

Category 37, 97 words.
trip, journey
holiday, vacation
transportation
automobile
shipping, ship
airplane

Category 38, 175 words.
country, nation
government
monarchy (king)
election, politics
congress, legislature
revolution, revolt

Category 39, 75 words.
law
court
crime
prison

Category 40, 118 words.
economy
money
finance
bank

Category 41, 143 words.
war
campaign, victory
weapon, arm
Army, Navy, Air Force
officer
military unit
naval ship

Category 42, 153 words.
thing, object
matter, thing, affair
manner, way
order, disorder
quality, characteristic
cause, effect

Category 43, 115 words.
part, particle
top, bottom, side
point, dot
space, place
direction

Category 44, 55 words.
number
quantity
measurement

Category 45, 80 words.
act, deed
renewal
motion
activity

Category 46, 35 words.
phenomenon
sound, noise
radiance, glow

Table 2

Sample Topic: 04 WEATHER 28 words.

Logical Organization.	Alphabetical Organization.
1.4 weather	breath of wind
4.0 climate	breeze
1.9 cold, coldness	climate
3.2 cool, coolness	cold, coldness
1.5 heat, warmth	cool, coolness
3.0 temperature	dew
6.6 thermometer	fog
2.0 rain	frost
2.5 shower (rain)	gust of wind
5.8 rainbow	hail
2.0 snow	heat, warmth
7.8 snowflake	ice
2.4 ice	lightning
1.4 wind	mist
5.2 breeze	rain
6.3 breath of wind	rainbow
5.7 gust of wind	roll of thunder
3.6 thunder	shower (rain)
4.1 thunderbolt	snow
4.3 roll of thunder	snowflake
3.3 lightning	storm
5.3 hail	temperature
3.1 fog	thermometer
3.1 mist	thunder
3.9 dew	thunderbolt
5.0 frost	weather
1.6 storm	whirlwind
6.2 whirlwind	wind
4.9 hurricane *	

---

* 'hurricane' has been added from the AHWFB. Its FPM is 13.424, giving it an equivalent Eaton ranking of 4.9. Some additional weather terms in the AHWFB were evaluated, but did not merit inclusion because of frequencies below 7,000: 'cyclone' (FPM 2.6203, ranking 10,900) and 'blizzard' (FPM 3.5043, ranking 9,400) were two of the words evaluated.

Table 3

A list of forms that include two or more different concepts or objects in their meaning extensions. A semantic link is present in all pairings. This is only a partial listing of polysemantic word-forms.

FORM	CONCEPTS
apartment	1. an individual dwelling for a family. 2. a building containing many individual dwellings.
appointment	1. a date, a planned meeting 2. a post, job, or position.
balloon	1. small--a child's toy. 2. large--an airship, a lighter-than-air transport.
bank	1. mound of earth by a river. 2. an institution for the storage of money. 3. a row of similar objects (e.g. a bank of batteries).
board	1. wooden plank. 2. food and drink.
cabinet	1. storage compartment or small closet. 2. governmental advisory group of ministers.
cardinal	1. a church cleric or dignitary. 2. a red bird.
cell	1. a basic organizational unit (e.g. of a political group). 2. the smallest structure of a biological organism. 3. a room in a prison.
chest	1. part of the body. 2. a box or trunk.
class	1. a grouping or collection of items or people. 2. a group of students.
coal	1. a mineral used for burning. 2. any glowing ember.
conception	1. an idea, concept, or image. 2. the beginning of life for an embryo.

Table 3 (contd.)

FORM	CONCEPTS
correspondence	1. similarity, matching of two objects. 2. an exchange of letters.
date	1. a numbered day on the calendar. 2. a meeting or appointment.
demonstration	1. showing or explaining something. 2. a public protest or manifestation.
drop	1. the act of falling. 2. a small mass of liquid.
duty	1. obligation. 2. a customs fee.
engagement	1. a meeting or appointment. 2. a betrothal.
faculty	1. a specific mental ability. 2. the teaching staff of a school or university.
fall	1. the act of dropping. 2. the season (autumn).
gas	1. vapor. 2. gasoline.
horn	1. of an animal. 2. of a car. 3. a musical instrument.
industry	1. manufacturing. 2. diligence; exhibiting perseverance.
lock	1. a key-operated device. 2. a water-regulating gate in a canal. 3. a bend or twist of hair; a curl or ringlet.
lord	1. the Lord and Savior. 2. a titled nobleman.
memory	1. the ability or faculty of recall. 2. a specific remembrance of an event or happening.
office	1. a place of business. 2. a position, assignment, post, job, or working capacity.

Table 3 (contd.)

FORM	CONCEPTS
operation	1. any activity or undertaking. 2. a surgical procedure.
order	1. a fixed plan, system, or arrangement. 2. a state of peace and serenity. 3. an authoritative command. 4. a request to make or supply something.
painter	1. an artist. 2. a person who paints houses.
peak, summit	1. the top of a mountain. 2. any high point.
period	1. a length of time. 2. a dot or point.
pin	1. a sharp, pointed object similar to a needle. 2. a piece of jewelry.
pipe	1. a metal tube or conduit. 2. a device for smoking tobacco.
plant	1. vegetation that is growing. 2. a factory or industrial installation.
powder	1. a fine, dust-like substance. 2. cosmetic face-powder 3. gunpowder.
pupil	1. a schoolboy or schoolgirl. 2. the center of the eye.
review	1. looking over or examining something. 2. a magazine. 3. a military parade.
ring	1. a circle. 2. a finger decoration in the form of a circle. (KER 2) 3. a sound made by a bell. (SKER 3 and KER 2)
root	1. underground tendrils that form the base of a tree. 2. any source or foundation. 3. the stem of a word.
shade	1. a shadow protected from the sun. 2. a hue or tint of color, of meaning.

Table 3 (contd.)

FORM	CONCEPTS
sheet	1. of paper. 2. of bed.
shift	1. any change or alteration. 2. a group of workers.
solution	1. resolving a problem. 2. dissolving something in a liquid.
spring	1. a jump or leap upwards. 2. a coiled length of metal. 3. a season of the year. 4. a source of water in the earth.
square	1. a four-sided equilateral geometrical figure. 2. an area in a city bounded by streets on four sides.
staff	1. rod, stick, or pole. 2. a group of officers, workers, or assistants.
stream	1. a brook or small river. 2. the flow of people or things.
stroke	1. a blow or strike. 2. a blood clot in the brain.
tank	1. a large container for liquid. 2. an armored military vehicle carrying a cannon.
turn	1. a complete revolution or turn about an axis. 2. a change or reversal of a course or direction (to right, left) 3. a change in circumstances, policy, health, events, etc.
union	1. togetherness; the fact of being joined. 2. any unified or federated group. 3. a labor group or workers' confederation.
want	1. a lack, scarcity, or shortage. 2. a wish or desire.
watch	1. visual surveillance or guarding. 2. a timepiece worn on the wrist.
wave	1. an up and down signal with the hand. 2. the undulation of water. 3. a regular formation of hair.

Table 3 (contd.)

FORM	CONCEPTS
way	1. road, path. 2. method, manner.
work	1. labor; a job, a task. 2. a work of art, an oeuvre.
youth	1. a young person. 2. a time of life.

Note: The two or three concepts listed for each form are by no means the complete range of that form. Only a few of the more representative concepts are listed--those concepts whose foreign language translations have a high frequency. Words like 'order' and 'turn' have a great variety of dictionary definitions, and it is also difficult to draw a boundary between many of these meaning extensions.

Homonyms

This short sample of homonymic pairs shows concepts that are not semantically related, but that share the same word-form as a result of linguistic coincidence.

FORM	DISSIMILAR MEANINGS
ball	1. round object used as a toy or in sport. 2. a dance or dancing party.
bill	1. a statement of money owed; a note of paper currency. 2. the beak of a bird; the brim of a cap.
cape	1. a headland or promintory over water. 2. an article of outer clothing.
match	1. a similarity or correspondence. 2. a chemically-tipped piece of wood used to start a fire.
pawn	1. a piece of a chess set. 2. an item pledged to guarantee a loan.

Table 3 (contd.)

FORM	DISSIMILAR MEANINGS
post	1. a stake or pole. 2. a position, assignment, or job.
race	1. an ethnic classification. 2. a competetive speed run.
rest	1. that part which is left over or remains from something. 2. relaxation, repose, or refreshment.
swallow	1. a gulping motion in the throat. 2. a bird.
temple	1. a place of worship. 2. the side of the forehead.

It is inevitable that such coincidences will occur, since language has a relatively small inventory of words (ca. 30,000 to 50,000) to express millions of different concepts and objects of reality. Given the fact that the average length of a word is between two and seven sounds (or letters), and the fact that most languages have a sound inventory of 30 to 45 sounds, it is not surprising that word-forms have a great burden in concept-expression and concept-representation.

Table 4

Synonym Pairings

In examining this list it must be remembered that no two words can substitute for one another in all environments and style uses. This list is not complete. It is arranged according to TOPICS and includes only noun examples in the topical vocabulary checklist.

1. WORLD
sod - turf
dock - pier
coast - shore

3. CITY
store - shop
factory - plant

4. WEATHER
heat - warmth
fog - mist

5. ELEMENT, MINERAL
china - porcelain
hue - tint, shade

6. ANIMAL
donkey - ass

7. BIRD
fowl - poultry
feather - plume
pigeon - dove

10. PLANT
shoot - sprout

11. TIME
epoch - age

12. PROFESSION
profession - occupation
office - bureau
personnel - staff
doctor - physician
lawyer - attorney

13. ADMINISTRATION
advice - counsel
help - aid, assistance
obstacle - barrier
order - command

14. PERSON
foreigner - alien

17. FOOD
gravy - sauce

18. HOUSE
cellar - basement
foyer - entrance hall, vestibule
hall - corridor
pillar - column
couch - sofa, divan

19. MACHINE, TOOL, CONTAINER
motor - engine
shovel - spade
string - cord
plank - board
bag - sack
pail - bucket

20. CLOTHING
pants - trousers

21. BODY
stomach - belly

22. SICKNESS
ache - pain
pill - tablet

Table 4 (contd.)

23. LUCK, CHANCE
probability - chance
fate - destiny
misfortune - hardship
agony - anguish, misery

24. MIND, INTELLECT
meaning - significance
concept - notion
understanding - comprehension
ability - aptitude

25. FEELING, SENSE, EMOTION
(feeling - sensation
(feeling - emotion
bliss - rapture, ecstacy
longing - yearning
(gift - talent
(gift - present
anger - wrath
fury - rage
zeal - fervor, ardor
pity - compassion
(comfort - consolation
(comfort - luxury, ease
(a worry - a care
(worry - anxiety
faith - belief
trust - confidence
respect - esteem
boredom - monotony
laziness - sloth
courage - bravery

26. EDUCATION
test - examination
mistake - error

31. COMMUNICATION
argument - quarrel
answer - response

32. WRITING
author - writer

34. COMPANY, SOCIETY
pardon - forgiveness

35. SOCIAL
feast - banquet
ball - dance
donation - contribution

36. SPORT
obstacle - hurdle

37. TRIP, JOURNEY
trip - journey
traveller - tourist
vacation - holiday
automobile - car

38. COUNTRY, NATION
border - frontier
kingdom - empire
monarch - sovereign
earl - count, baron

39. LAW
law - statute
fine - penalty
prisoner - convict
prison - jail
thief - robber
will - testament

40. ECONOMY, BUSINESS
company - firm, corporation
trade - commerce
value - worth
salary - wage, earnings

41. WAR, MILITARY
attack - assault
spoils - booty
exile - banishment
refuge - haven, asylum
spear - lance

42. THING, OBJECT
(material - matter
(material - cloth
thing - object
event - incident
conduct - behavior

Table 4 (contd.)

(state	- condition	43. PART, POSITION	
(state	- nation	bit	- piece
kind	- sort	end	- conclusion, finish
contrary	- opposite	start	- beginning
defect	- flaw	middle	- center
evil	- wickedness	rest	- remainder
safety	- security		
goal	- aim, purpose	44. NUMBER, QUANTITY	
		amount	- quantity
		shortage	- lack, want
		45. ACT, DEED	
		act	- deed
		turn	- revolution, revolving

These informal concept-pairings have been established to obtain accurate frequency numbers for the total concept. There is no implication that the two members of any synonym pair are identical in meaning. The bracketed pairs ((indicate two synonym pairs together with a concept difference, and the synonym pairs are necessary to establish the polysematic concept difference.

This list is limited to nouns. Verbs and adjectives also provide numerous example of one concept being expressed by two word-forms.

Even though many synonym pairs are rather loose ('kingdom - empire, spear - lance, traveller - tourist'), it is felt that a better frequency indication is obtained for the total concept rather than for each of the two individual word-forms.